AU/ACSC/2016

AIR COMMAND AND STAFF COLLEGE AIR UNIVERSITY

STRUGGLING WITH "EXCELLENCE IN ALL WE DO":

Is the Lure of New Technology Affecting How We Process Our Members' Information?

by

Raul F. Martinez Jr., Maj., USAF

A Research Report Submitted to the Faculty

in Partial Fulfillment of the Graduation Requirements

Advisor: Dr. Ed Ouellette

Maxwell Air Force Base, Alabama
Feb 2016

Disclaimer

The views expressed in this academic research paper are those of the author and do not reflect the official policy or position of the US government or the Department of Defense. In accordance with Air Force Instruction 51-303, it is not copyrighted, but is the property of the United States government.

TABLE OF CONTENTS

Abstract

Inaccurate Department of Defense member information poses a significant threat to mission accomplishment. Natural disasters and threats to homeland security have amplified the need to account for military members and their families. Unfortunately, the way members' information is managed today is far too complex and riddled with risk. Why is a members' information duplicated across multiple disparate databases? To better assure the military is prepared before disasters strike, the Air Force should minimize duplicate data fields across multiple Military Personnel databases. The purpose of this paper is to provide a viable solution within a given set of constrains that the Air Force can implement.

Utilizing the problem solution method, this paper identified gaps such as multiple systems having duplicate data fields, databases not able to talk to each other, and continuously changing information. After evaluating multiple alternatives, it was determined the Air Force should combine technologies that are currently in use in the civilian sector. Utilizing the best of multiple technologies would reduce the risk of implementing only one alternative, ultimately leading to a viable integrated solution that will be considered the real source of truth. This single source of truth will maximize auditability, reduce cost through the phasing out of legacy systems, reduce keying errors and improve data confidence to ultimately improve mission capability.

Section 1: Introduction

On August 31, 2005, the call went out, "Commanders, the city of New Orleans has taken a direct hit from Hurricane Katrina and the state of Louisiana needs us; recall your people." Performing a recall of all members seemed like such an easy task, or was it? The 149th Air Guard unit knew it had 24 hours (24 for mobilization and 28 for activation of AFRC and ANG units)[1] to report and be completely ready to deploy, but it seems like 24 hours is not enough time when member information is not correct. Accountability is more than just making contact with personnel responsible to accomplish a mission. It is being able to provide effective communication in the event of a catastrophe, both up and down the chain of command, and being able to maximize the amount of time a member has to report, resulting in a more focused, less stressed Airman.

Natural Disasters such as wildfires, floods, tornadoes, and hurricanes affect thousands of people every year. Terrorist attacks have long been a threat to the national security of nations. Whether natural or terroristic, these threats can bring substantial injury, loss of life, destruction of property and large-scale displacement of large numbers of people. Treating these incidents in discrete phases – before, during, after – will enable an analysis of the accuracy of the data by focusing on the importance of accurate data at the correct time of the process.

In the first phase, "before an incident," preparedness is the key to success. Accurate personnel accountability information, when it is needed, will help to improve the rate of success during and after an incident has happened. Unfortunately, the problem with inaccurate personnel accountability information is not realized until after an incident has taken place and this, the "during" and "after" phases, is when essential operations are needed most. According to

Pentagon Inspectors, “Many key U.S. Air Force bases and organizations never adequately created or tested plans about how to continue essential operations in case of a terrorist attack or natural disasters from earthquakes to tornadoes.”[2] Plans frequently reflect actions taken during an incident such as, “what if we have bad data and cannot make 100% contact, what do we do?”

Minimizing the opportunity for failure before it happens should be the focus. Another way to ask the question is, “looking at the process holistically, where does it start, what are the potentials for error and what are we doing to mitigate those risks?” Starting with accurate data is the best way to minimize the problem of “garbage in, garbage out.” Is the process of ensuring accurate data so difficult that the Air Force is willing to accept less than perfect accuracy? Why is a members’ information duplicated across multiple disparate databases? To better assure the military is prepared before disasters strike, the Air Force should minimize duplicate data fields across multiple Military Personnel databases.

Section 2: Background

Military databases have three things in common that all contribute to the potential for problems with accurate information which are multiple systems having duplicate fields, databases not able to talk to each other, and continuously changing information. While very little can be done about information changing in a person’s life, the military can do something about how it manages its members information.

Technology is growing at an incredible rate. “Forty years ago, Intel's first microprocessor had 2,300 transistors; today's microprocessors have over 2 billion transistors.”[3] With the rate of progress that has been made in technology, it has been difficult to have the foresight to know what to retire and when. While the retirement of systems that are still working may not be the

answer or even fiscally responsible, there should be an analysis of how the Air Force is systematically managing their most critical assets: Their people and information. Managing the accuracy of their members' information has been a challenge. Many potential points of failure include people, process, technology, and information. For example, the Air Force relies on their members to provide timely and accurate data. While members may provide accurate data, the act of keying the same information into multiple databases, also known as double or multiple entry, inherently increases the risk of error. Studies have been conducted analyzing data entered across two databases, and the results did not fare well. For example, "data in clinical research databases was analyzed for external inconsistencies. The analysis consisted of 1,006 patient records that were incidentally entered in two different databases at the same time. Furthermore, the analysis evaluated discrepancies between the records of the same patients in the two databases in the following fields: medical record number (MRN), date of birth (DOB), first and last name, number of treatment sessions, and the dates of the first and last treatment session. All of the demographic information fields were entered on one screen in both databases, and all of the information related to treatment was entered on another screen." [4] An analysis of data errors in clinical research databases found that errors in the data were common, including incorrect and missing information. "Error rates detected by the double-entry method were as high as 26.9 percent corresponding to a 13.5 percent error rate in each of the databases. Errors were due to both mistakes in data entry and misinterpretation of the information in the original documents."[5]

In addition to having multiple duplicate database fields, military personnel databases are unable to interface with each other. Whether this is intentional because of a perceived vulnerability or just an inherited flaw due to different designers, data systems continue to operate in silos. Problems with merging databases seem to be never ending. For example, database

models may be different. Databases usually fall into three different categories, federated, relational, and object-oriented.

In a federated database, multiple databases seem to function as a single entity. Having a federated database works well when you have large amounts of data and want to minimize the retrieval time of certain information. For example, in a bank, a federated database model works well when information is broken up by year. The reason it works well is because queries of information are usually based on time, such as; show me all of my deposits in the past six months. However, if the information is broken out by customer ID the time to query would be considerably longer. This is because a given set of transactions will have a seemingly random or a Poisson distribution, which means the system will have to review all of the records as opposed to a pre-segmented set of information.

Originally proposed in the 70's, relational databases utilize tables that have similar attributes and are linked together through logical relationships. Additionally, further analysis of information is performed through the use of Structured Query Language (SQL) that communicates between tables and makes things such as queries or reports happen. The availability of programs such as Microsoft Access has increased the familiarity of relational databases thereby increasing their popularity for small-scale database design primarily because of its low cost and no help from IT is required.

Another category is an object-oriented database. An object-oriented database is one that accesses objects directly with no need for a query sublanguage such as structured query language (SQL) or utilizes a call level interface such as Open Database Connectivity (ODBC). ODBC is the standard that allows databases to talk to other databases or tables-like structures such as Excel. Conversely, object-oriented databases utilize objects then use code to modify or replicate

objects with the system. One advantage of object-oriented databases has been both the objects and the code will use the same model representation which means there is more consistency in this environment. While a little more complex than a relational database, object-oriented databases also have the ability to handle much larger volumes of data. On the other hand, a relational database is based on a relational model or a grouping of tables that have the same attributes. This type of database uses SQL for querying records. The advantage of relational databases is their simplicity and ease of setup. Unfortunately, this ease of setup also allows for multiple variations in design. Introducing all of these variables into the decisions made during the design process further supports the fact of differences in database design.

Initial development of databases could also play a factor. Unless intentionally designed to work with other systems, database designers will optimize on what best fits the customers' requirements. As the example above shows, database designers optimized queries and information to maximize the productivity of their systems. In most cases, databases are not designed for integration with other systems. According to a presentation at Yildiz Technical University, database design will take three steps to design, the conceptual, the logical and the physical. The conceptual step is the highest level of design. It includes the analysis of data types, relationships, and constraints. The next step is the logical step. In this step, the designer is going through the implementation of the conceptual model. This is also where decisions are made to determine the optimal type of database to be used. The last step in the process is the physical. In this step, key considerations of memory, storage, indexing and management of the system will take place.[6]

Even if two databases are in the same category, such as both relational, it does not mean the data schemas or the way they describe the data is the same. To describe data in a database

codification criterion is used. This criterion describes each data point through the use of attributes initially assigned. For example, data can be grouped and described by assigning a field as text, a number, or date. The advantage of doing this allows data to be formatted and standardized across the system. Unfortunately, the seemingly endless combinations of data schemas that could appear in a single database introduce a new level of complexity to integrating systems.

Technology has produced new capabilities and matured others; unfortunately, we have not been very effective in phasing out systems that duplicate capabilities. Maintenance and storage solutions are expensive. While the cost of technology may be decreasing, the cost to manage continuously growing information is increasing.[7] A study done at Carnegie Melon University states the amount of storage sold is expected to sustain an annual growth rate of 60 percent per year. Conversely, this growth rate is accompanied by a 50 percent decrease in the cost per byte of storage in large scale computing.[8]

Having fewer data fields and databases to manage reduces the impact when our member or their family's information changes. Additionally, when changes are required, ensuring all of the impacted fields are updated accurately is difficult. Member's information is constantly changing. For example, as of 2006, active duty personnel have a Permanent Change of Station (PCS) every 48 months.[9] Other than moving, changes to a member's information could include getting married, divorced, having or adopting a child and even changing a phone number. The process for entering data and understanding which system to complete the transaction has been a challenge. Duplicate data fields between different personnel management information systems are at higher risk of typographical errors due to human keying errors, missed fields, and not

having an understanding of the significance of the data captured especially if the same information was already recorded in other systems.

Considering the already identified gaps in the current state, this paper will employ the problem solution method. A problem is defined as "a perceived gap between the existing state and the desired state."[10] Clearly describing the existing state and the desired state will set up the criteria for options to close the gap. Analysis of options from multiple perspectives such as the user, the taxpayer, and program owner will ultimately reduce errors, duplication of effort, align to industry standards and ultimately could potentially save taxpayers money.

There may be an opportunity to utilize existing processes in new ways to validate information such as utilizing the recall roster that is prepopulated with information from a single repository then manually passed around to verify information. In the event there is an error, the correction should be made in the Defense Enrollment and Eligibility Reporting System (DEERS) as opposed to the recall roster, which would lead to improved data accuracy.

The central point of this research would require an evaluation of options to minimize redundant database fields. The intent is to improve the accuracy of data in our military personnel systems. It will require the evaluation of past attempts to integrate massive amounts of data from multiple systems as well as options available to the civilian sector. Other criteria to consider are identifying capabilities needed that will minimize errors and how the Air Force is defining a successful project. A rigorous literature review of this topic will also be conducted to ensure this topic has not been evaluated. Criteria for evaluation will have to be clearly spelled out to ensure repeatability. For example, what are all of the systems involved in the scope of this project? What are some of the common fields and their format? What are the differences in the acquisition guidance given to each system? What type of culture or company has already

completed this type of task and what can we learn from it? What projects or companies have failed at this task and what can the Air Force learn from it? The proposal could help improve the accuracy of member data and improve accountability. Additionally, standardized information and processes will help reduce confusion for our members and those that maintain members' information.

Section 3: Overview of Military Personnel Mass Storage Systems

Failed database integration attempts have cost taxpayers billions. Continuing to fund database systems in silos without an integration plan may be worse. Siloed database systems not only show fiscal irresponsibility but waste valuable time, resources, and increase the risk of vulnerability. Non-integrated databases force actions that create duplication of effort and multiple potential points of failure resulting in loss of accuracy of information. In a recent article, Harvard Business Review talks about data's credibility problem noting, "knowledge workers waste up to 50% of time hunting for data, identifying and correcting errors, and seeking confirmatory sources for data they do not trust. Moreover, consider the impact of the many errors that do leak through such as an incorrect laboratory measurement in a hospital can kill a patient. An unclear product specification can add millions of dollars in manufacturing costs. An inaccurate financial report can turn even the best investment sour." [11] In the event of an emergency, the accuracy of information is critical in ensuring proper contact is made. For example, during the 149 FW's last recall exercise, failure to contact almost 9 percent of 620 members resulted in unaccounted for members, which lead to additional work.[12] During the review, there were significant discrepancies when data in recall rosters was compared to the data in the Air Force Personnel Accountability and Assessment system (AFPAAS). Some argue that it

is the responsibility of each member to maintain the accuracy of their data. While it is true, the accuracy of information is the responsibility of the data owner; integrated technology and improved processes can help improve data integrity.

The Air Force has six authorized databases that make up the Air Force Directory Services (AFDS) platform. AFDS provides identity management capability for systems such as Global Combat Support System (GCSS), Global Command and Control System (GCCS), AF Global Address List, and Active Directory.[13] It is important to note, each of the six authorized databases is an autonomous system that was not designed to interact actively with other systems, meaning that in some cases data can be extracted but not uploaded.

Defense Manpower Data Center (DMDC) "maintains the largest, most comprehensive central repository of personnel, manpower, casualty, pay, entitlement, personnel security, person identity and attributes, survey, testing, training, and financial data in the Department of Defense (DoD)."[14] It serves under the Office of the Secretary of Defense (OUSD) to collate personnel, manpower, training, financial, and other data for the Department of Defense (DoD). This data catalogs the history of staff in the military and their family for purposes of healthcare, retirement funding, and other administrative needs."[15] Their mission is to be, "the DoD's source for enterprise human resource information, providing secure services and solutions to support the Department's mission."[16] With 35 million personnel records, the DMDC is at the top of the list of mass storage centers for the military, capable of performing 5 million transactions per day to verify identity, benefits, and entitlements.[17] Feeding into the Person Data Repository (PDR), also known as DEERS, is Real-Time Automated Personnel Identification System (RAPIDS), Defense Biometric Identification Data System (DBIDS) and Trusted Associate Sponsorship System (TASS). This last system is most known for establishing TRICARE benefits eligibility making

this system incredibly important given the requirements of the Affordable Care Act. The Patient Protection and Affordable Care Act (ACA) Employer Mandate / Employer Penalty, originally set to begin in 2014, was delayed until 2015 / 2016. ACA's "employer mandate" is a requirement that all businesses with 50 or more full-time equivalent employees (FTE) provide health insurance to at least 95% of their full-time employees and dependents up to age 26, or pay a fee of $2,000 per full-time employee per month starting in 2016.[18]

Air Force Global Address List is used to maintain fax, phone, mobile phone, and address information. Additionally, job title, department, or company information will be stored in this system. While this seems to be a module within DMDC, access to information is through the MILConnect portal via Common Access Card (CAC) and other DMDC modules are not available to the user.

Military Personnel Data System (MilPDS) is the "primary records database for personnel data and actions that occur throughout every total force Airman's career. The system is also used to initiate Airman pay actions, maintain Air Force accountability and strength data and support a host of interactions with other Air Force processes and systems that rely on personnel data."[19] According to AFI 36-2134, "MilPDS shall be used to update and maintain Strength Accounting Duty Status Program Reporting (SADSP). The SADSP exists to enhance total force accountability and improve crisis response."[20] Additionally, strength accounting of each member's duty status potentially affects funding of personnel and is used to manage the Operations Tempo (OPSTEMPO) down to an Air Force Specialty Code (AFSC) level.

Manpower Programming and Execution System (MPES) is used to program manpower as a part of the Annual Air Force Budget that is approved by The Congress. According to AFPD 38-2, "The Congress controls manpower levels by authorizing and funding military end

strengths, funding the civilian work force, establishing military grade distributions and directing human capital resources and programs through legislation each year."[21] By focusing on the end strength of military and civilian personnel, resources are provided to support approved force structure and missions. It is important to note, end strength does not drive mission changes; rather it is a way to ensure the right resources are available for assigned missions.

Advanced Distributed Learning System (ADLS) is a system that brings together 16 site partners to provide a wide variety of training to the military member. Tailored for new and future warriors, ADLS provides new training opportunities in a faster and more consistent platform. Operating on the Global Content Delivery Service (GCDS), ADLS can leverage commercial Internet technology to accelerate and secure DoD Web content and applications across the NIPRNet (non-secure network), SIPRNet (secure network), and CX-SWA (coalition network) 24x7.[22]

Defense Civilian Personnel Data System (DCPDS) also known as "CIVMOD" is primarily used to store employment information. Employment verification is a feature that is provided to current DOD employees allowing them to send employment and salary information to an external organization via a web-based tool called MyBiz+. MyBiz+ is a self-service, web-based, employment verification system used to show proof of employment to an external organization. Directly connected to DCPDS, a DOD employee can show proof of employment, salary information and other details pertaining to employment with the Department of Defense.

It is true that failed database integration attempts have cost taxpayers billions and continuing to fund database systems in silos without an integration plan has had a negative impact on the Air Force's ability to do its job. Focusing on key issues, as opposed to all issues,

will help to focus efforts on the most significant pain points. In this paper, key issues are defined as opportunities that have the potential to significantly impact the success of an objective.

Section 4: Key Issues

When companies merge in the civilian sector one of the top priorities is to analyze the personnel information across both companies. Consolidating multiple systems not only reduces costs through reduced headcount and resources but it allows for better Information Technology practices. As opposed to two merging companies, the Air Force can be considered one company. With that said, the need to analyze how personnel information systems are utilized has not been a priority. While there is no set standard of when database systems should be evaluated holistically, it is wise to take this opportunity to analyze the systems that store military personnel records. More importantly, is it time to evaluate personnel systems across our sister services? The military, as a whole, is migrating towards utilizing multiple services to accomplish a mission; this is also known as a joint environment. The accountability of personnel, in theater, should not be siloed by service; rather, it should be shared across services to enhance productivity in meeting objectives. In essence, this is the role of JOPES (Joint Operations Planning and Execution System). JOPES is the "integrated, joint, conventional command and control system used by the Joint Planning and Execution Community (JPEC) to conduct joint planning, execution, and monitoring activities." [23] Utilizing this system as a benchmark will help in developing and proposing a feasible option.

Personally Identifiable Information (PII) type fields are frequently used to identify and retrieve user's records. The problem is how this information is stored, more specifically; classification of this information may be different across platforms. Another issue is maintaining

the same information across multiple systems. While the examples above are designed to focus on different things, there may be advantages to integrating these systems.

One factor that may prove to be a problem is the various owners or program managers of the various data systems. For example, the Navy oversees DCPDS while DMDC reports directly to Personnel and Readiness under the Secretary of Defense. Another agency, Air Force Personnel Center (AFPC) runs MilPDS. The problem with having various owners can be translated into various funding streams. Congress appropriates money in categories. Once a project has been funded or appropriated to a specific category, it is illegal to utilize money from one project to supplement another. This is the reason most systems are designed the way they are; but it is time to analyze our ecosystem and determine whether, with current technology, the risk of integration exceeds the impact of problematic data fields.

Integrating heterogeneous data sources is taking new forms. The once popular Enterprise Resource Planning (ERP) is being replaced with virtual master databases utilizing wrapper code to translate various schemas. Often used in translating a schema into one that a host system will accept, wrapper code is a thin layer of code that is used to translate information. For example, if a data field is set to text in one system and the host needs it to be a number with a certain set of decimal places, wrapper code helps in translating from one system to the next. Data warehouses are also a popular choice, but some drawbacks include the time it takes to run various queries and resolving semantic conflicts between disparate data sources.

Attempts to integrate information across multiple systems in the military have been done. Unfortunately, unsuccessful attempts cost the taxpayers billions of dollars. While these projects are not intentionally designed to fail, it is important to note that it does happen. For example, an article on DSS Resources, states, “Database centered projects for decision support and

transaction processing do fail. How often? According to a study of IT projects by The Standish Group reported in 1995, ‘Only 9% of projects in large companies were successful. At 16.2% and 28% respectively, medium and small companies were somewhat more successful.’”[24] In 2005, a U.S. Air Force IT modernization project consisted of utilizing an ERP system to update the Air Force’s logistics systems with Oracle software. The project was called the Expeditionary Combat Support System (ECSS). According to Shaw, "the financial decision to cancel the project was made when Air Force leadership determined that another billion dollars and eight more years would produce one-quarter of the planned system capabilities." [25]

Section 5: Potential Solutions

Potential solutions or options should be varied and imaginative. Ideas should not be limited by self-imposed constraints. Minimizing potential solutions will be done using an evaluation criterion. Criterion such as: How structured is the data that the Air Force is working with and what is the impact to their ability to meet the objective of minimized duplicated fields across database systems? Having some form of structure is important, whether it is achieved with predefined schema or is translated into a system that can read it after the fact. The goal remains the same; develop something that will allow for a high level of record accuracy. What is the standard around version control and how is the Air Force managing one source of truth? It is important to understand, the more individuals or inputs to a system of record has the greater chance of variability. In today’s situation, the potential for error is six times more likely for each similar field. In other words, the Air Force has accepted an inherent flaw in the way data is managed, which could lead to a significant risk of failure.

Before considering any solution, an evaluation of People, Process, Technology, and Information (PPTI) must happen. From the process perspective, it is important to evaluate all of the impacted business processes, have a change management plan and have strong committed leadership. A business process is defined as a collection of linked tasks, which find their end in the delivery of a product or service to a client.[26] In addition to understanding what and how business processes are accomplished, it is important to identify and manage the risk associated with each process. Evaluating business processes aids in the reduction of non-value added tasks, which could improve the success of the overall project by automating only tasks that add value.

Change management helps to transition organizations through the complexities of a significant project to ensure it meets its intended outcomes. Utilizing an industry accepted approach, a change manager will guide individuals within organizations through three phases: "preparing for change, managing change, and reinforcing change."[27] An effective change manager will understand the target audience and will minimize confusion through carefully timed effective communication. They also gather and manage feedback through corrective action plans to ensure adaptation of changes.

Having a strong, committed leadership means one that understands the vision and is willing to stand behind the plan to link strategy to execution. In an article written for Harvard Business Review, Paul Leinwand states, "only 8% of leaders are good at both strategy and execution."[28] One of these leaders is Starbucks CEO Howard Schultz. He believes by becoming the architect of the capabilities you need, the chief of builders, you can effectively translate the strategic into the everyday.[29]

It also means listening to concerns, evaluating risk and knowing what actions to take to minimize the potential for failure. The advantage of having a strong, committed leadership is

having the ability to focus on what matters most as opposed to aligning resources to areas that are not in alignment with objectives.

After evaluating processes and who should be involved, it is important to analyze from an information perspective. In the databases being analyzed, data fields such as name, rank, address, phone, and social security number are each repeated in each of the six databases. While this may not seem like a big problem, each member has to manually update the same changes in six locations. If even one field were missed, it would be considered a defect with the potential to have serious consequences.

Additionally, having a strong, committed leadership means choosing the right people who WILL look at all the angles before any decision is made. Evaluation of alternatives will help in determining the best solution going forward. Additionally, evaluating potential problems for each of these solutions will address the risk associated with the decision. Balancing the benefit and risk associated with each decision will ultimately lead to a better decision. Furthermore, after evaluating the benefits of an alternative and establishing/deciding a clear winner, it may not make sense to choose that option because it may involve significant risk. For example, purchasing a technology solution that does everything needed may seem like a good idea but if the subsequent modifications or system upgrades – and their inherent costs – are not accounted for it may be a bad decision.

Lastly, an evaluation of technology is required. This evaluation of various technologies available to meet the objective will evaluate the following possible solutions: leave as is, use an integrated system such as ERP, utilize NoSQL or use data hubs. The following is an overview of each of the proposed systems or technologies that are readily available on the commercial market today.

Leave as is

One option is to leave the system as is. In most cases, our current systems are designed as relational databases. MilPDS has been developed to resemble a federated database system with the exception of being one system as opposed to multiple systems. The advantage is having only one data source. Unfortunately, this single data source works independently of other databases. This lack of communication results in a disparate database that has limited capability in accepting automated information. While this system has the ability to feed other systems, it will not accept information from any other system, which forces users to manually enter information. It was mentioned MilPDS resembled a federated database but not quite. Remember, a federated system is designed in such a way that multiple systems all work together to function as one. In this case, MilPDS has several tables that theoretically could represent multiple systems. For example, there are three different master tables (Active Duty, Air Reserve and Air Guard) that can act like three different databases. In each of these tables, records are further broken out by officer and enlisted as an example of the relational database. The advantage is the speed of execution. Sorting through a fraction of 500,000 active records increases process velocity and decreases the speed to retrieve a record. The problem with this model is in when a member transfers from Active Duty to Guard or Reserve or vice versa and transferring members from one service to another is not any easier. In most cases, the transfer of information is a manual process that takes information from one system and enters it into another. Optimally, this process would be simplified if only a modification to an object's metadata was required.

While this option is not optimal, the system is currently working. However, having multiple disparate systems is extremely expensive and labor intensive and as noted above subject to much higher risk.

Integrated systems and ERP

Originally designed as a repository for manufacturing information, ERP has evolved to other areas of organizations. As a single repository, it has the ability to grow in capability by adding modules that specialize in specific functions. For example, Human Resources and Finance have been two modules that have proved to be very successful in the civilian sector. Oracle states on their site how, "Silicon Valley YMCA CFO, Ed Barrantes, shares how Oracle ERP, EPM, and Sales Cloud will help double its subscriber base and achieve financial excellence."[30] Just like tracking inventory, procurement of assets, or even production information, ERP systems can track information related to the different events that happen in an employee's life. ERP is evolving.

Starting with the same basic concept, ERP has evolved to ERP II with the addition of web services and the ability to interact with other systems. Historically, the use of middleware to communicate between computers has been difficult due to complexity of the code required to make communication happen. The use of a web service makes this process very easy. A web service utilizes standardized code to create something similar to a portal to other systems while controlling access resulting in more secure information with the repository. Utilization of this standard has allowed the expansion and collaboration of information across multiple platforms. For example, "JPMorgan, a leader in investment banking, asset management, private equity, custody and transaction services, middle market financial services, and e-finance, uses web services to connect Excel spreadsheets to UNIX-based financial data."[31] In this case, the information was heterogeneous. While the environments or systems are different, heterogeneous means the computer ecosystem is within a single company. Conversely, Con-Way, a $2 billion

transportation company based in Ann Arbor, Michigan must interact with many suppliers. They use web services to share electronic shipping data.[32]

ERP has had success in smaller companies but has shown a history of failing in extremely large-scale deployment. According to researchers at the Miter Corporation, the problems that Department of Defense ERP experienced have been with large numbers of interfaces that add both complexity and risk to the programs.[33] Complexity equals cost. Increasing the number of interfaces also increases the potential points of failure that must be maintained as well as manpower needed to ensure operability of such interfaces

NOSQL

NoSQL originally, non-SQL or nonrelational databases provide storage and retrieval of information but in a nonrelational way. However, this does not mean they will not interact with relational databases. NoSQL databases will have to interact differently. For example, in a relational database, an SQL query has the ability to combine records from multiple tables, an action also known as join queries. This query then produces a new record with data that was joined from separate tables. In the NoSQL, the same output can be achieved but with many more queries. Since the NoSQL database is significantly faster, the additional queries have no impact on comparative performance. Additionally, NoSQL databases have the ability to interact with existing systems, our analysis can expand to other opportunities such as scalability and performance. According to a MongoDB whitepaper, “Relational databases were not designed to cope with the scale and agility challenges that face modern applications, nor were they built to take advantage of the commodity storage and processing power available today.”[34] While these database types may be new, the need to handle new multi-structured data types or scale beyond

the capacity constraints of existing systems is not. Motivation has been driven by the desire to identify viable alternatives to expensive proprietary database software and hardware. [35]

Data Hubs

Finally, data hubs do more than store data in one place. One significant feature of a data hub is the ability to reduce duplicate data fields. Duplicate data fields have the potential to increase variability over similar record fields. This variability leads to mistrust of information, which in some cases, creates the need to have redundant verification process steps added to help validate information. In worst cases, the data is abandoned altogether in favor of manually maintaining information in a user created data source such as the recall roster.

Another feature is the ability to work with very large amounts of data. The national weather center is currently utilizing data hubs to share and analyze information. For example, natural hazards such as Hurricane Joaquine produced massive amounts of data in various systems; data that can be used to evaluate and predict the effects of storm surges.[36] Given the amount of data produced, it is physically impossible to manually analyze information in a timely manner. Data hubs allows information to flow between systems freely thereby allowing computer systems to evaluate very large data sets at a much faster rate with an increased accuracy than ever before. Another example where data hubs could potentially prove useful is in the field of Intel. A vast amount of information is gathered yet it is physically impossible to turn all of it into intelligence. In this case, systems are needed to take on a larger share of the load.

Data hubs are made up of two integrated systems. The first system is the data set version control processor. The “Dataset Version Control System (DSVC), is a system for multi-version dataset management. DSVC’s goal is to provide a common substrate to enable data scientists to capture their modifications, minimize storage costs, use a declarative language to reason about

versions, identify differences between versions, and share datasets with other scientists."[37] The second is the version query processor also known as a data hub. The data hub is a "hosted platform built on top of DSVC, that not only supports richer interaction capabilities, but also provides novel tools for data cleaning, data search and integration, and data visualization tools."[38] As depicted in figure 1, data hubs require significant work to integrate all of the components correctly. However, the added value that is brought to the table may be worth it.

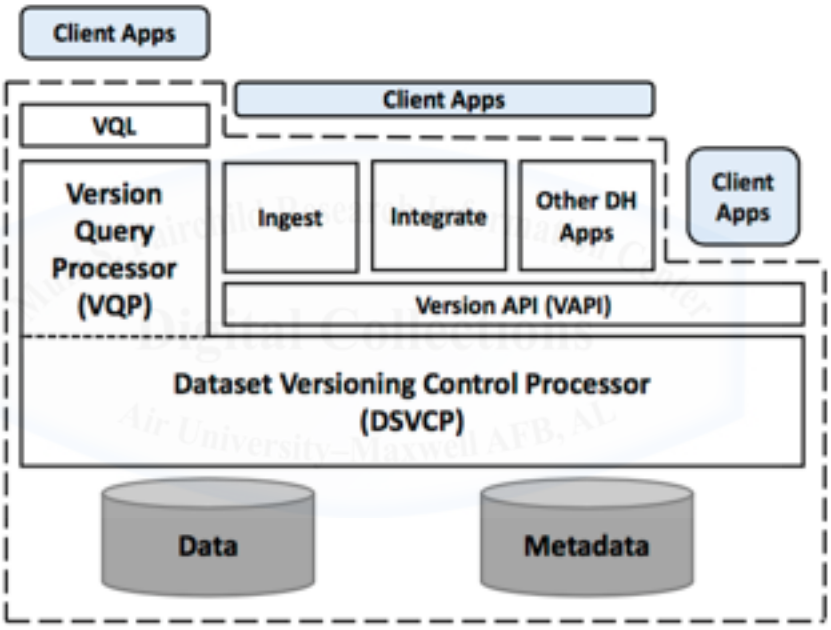

Figure 1: Data hub Components and Architecture

While data hubs seem like the easy answer, it is important to note data hubs do not promote the reduction of disparate database systems thus allowing legacy systems to continue. From a cost reduction perspective, this option is one of the more expensive. On the other hand, homogenizing data fields and serving data in and from different formats adds such significant value the risk may be outweighed.

Section 6: Solution Comparison/Evaluation

There are risks involved with every decision made. The key to an optimal solution is to maximize the benefit or value while holding risk to an acceptable level. It is one thing to simply maximize value; it is completely something else to maximize value while evaluating risk. For example, in the medical field, doctors must evaluate both the benefit and the risk associated with their decisions or potentially face a medical malpractice lawsuit. In medical decision-making (MDM), physicians must stratify the MDM into levels of complexity based on the nature and number of clinical problems, the amount, and complexity of the data reviewed by the physician and the risk of morbidity and mortality to the patient. After evaluating the information, the information is further evaluated in a matrix that definitively defines the level of complexity. After the level of complexity is defined, a point system is assigned to both problems and data reviewed. All three information points are evaluated against an MDM points table to quantify the risk as either minimal, low, moderate or high.[39] A similar process will be followed to help compare and evaluate proposed solutions. Lessons learned from past failed deliveries should also be considered in the analysis of systems going forward.

Evaluation Method

Used as a decision making model, the Pugh matrix is used to determine the best alternative from a group of alternatives. Developed by Stuart Pugh, who was a professor and head of the design division at the University of Strathclyde in Glasgow, the Pugh matrix evaluates alternatives against a baseline that is also known as the current state or the system that is currently in use.[40] As with other decision-making tools, the most important criteria are chosen and then utilized to evaluate each alternative. To further emphasize the importance of the criteria, each criterion is weighted to signify a greater level of importance. The alternatives are then

evaluated against the baseline. For example, if an alternative is about the same as the base line, the score of zero is assigned to that criterion. However, if the alternative is better then a score of one is given. Conversely, is the alternative is worse then a score of negative one is given. Instead of a simple three-point scale, a five-point scale will be used to show difference between better than and much better than. For example, a value of two will signify much better than, a value of one will signify better than, a value of zero will signify equal to, a value of negative one is worse than and finally a negative two is much worse than.

Decision criteria

Decision criteria are selected for this analysis to achieve three objectives. The first is maximizing the accuracy of information, the second is to minimize complexity from a users perspective, and the third is to minimize cost.

To maximize accuracy of information, a fully auditable solution must be achieved. The ability to aggregate information in various ways and at various levels to provide information such as the total number of individuals assigned, available for duty, status of training, and deployed to any given location without having to manually access several systems is critical. Additionally, the ability to validate an individual's pay, time and attendance, number of points and travel information, again in a single system, is just as important.

To minimize complexity from a users' perspective also known as operational complexity, an integrated solution is best. Minimize operational complexity is defined as reducing the number of systems a user must maintain. It is important to note it does not include the complexity of the system on the back end, or what a system administrator must do to maintain it. While providing a simple-to-use interface is nice, knowing there is a single source of truth is invaluable. Users should not be required to maintain the same information in multiple

repositories. Utilization of an Application Program Interface (API) that allows software applications to talk to each other or wrapper code, which translates dissimilar data types, is preferred to manual maintenance of information over disparate systems. Ultimately, a single location where multiple users can access information in a standardized format is optimal.

Finally, cost must also be minimized. To minimize cost, it is unacceptable to have a project last several years. While understanding all of the key components, testing, and validation is important, design to implementation time also known, as time to release should be minimized. Another way to address cost is to stipulate in the contract "any and all cost overruns will be the responsibility of the contractor." Unfortunately, this type of verbiage results in much higher-priced bids due to the significant risk the contractor has to take on.

Additionally, a plan to retire legacy systems will also go towards saving costs; both in management and maintenance of such systems. Management of legacy systems is expensive and there are no easy formulas to select which one to retire and when. A recent merger of a civilian firm resulted in the need for an analysis of such legacy systems. Their approach was to first analyze how much it was going to cost to maintain various platforms, especially those that were custom built. Platforms that needed specialized support were among the first to go.[41] Utilizing lessons learned from the civilian sector has proven to be most fruitful.

Evaluating these criteria in a Pugh matrix will help to identify the system that best meets the predefined objectives. While other decision tools are available, utilization of a Pugh matrix minimizes subjectivity by directly comparing alternatives to what is available today. In most cases, information used to determine why an alternative is better or worse than what is in place is readily available.

System Evaluation

After the criteria have been defined, the alternatives are evaluated against the baseline. The baseline is the system currently in place and the alternatives are: use an integrated system such as ERP, utilize NoSQL or use data hubs. Each is assessed against the baseline for each criterion and is given a value from the five-point scale. This value is then multiplied against the weighted value of each criterion to develop a score that can ultimately be compared. As shown in Table 1 below, the option Data Hubs has amassed the most number of points followed by NoSQL and finally ERP.

Criteria	Leave as-is	Weight	ERP		NoSQL		Data Hubs	
	(Baseline)		Scale	Points	Scale	Points	Scale	Points
Maximize Auditability	0	5	1	5	2	10	2	10
Maximize Interoperability	0	4	1	4	2	8	2	8
Minimize Operational Complexity	0	4	1	4	2	8	2	8
Minimize Time to Release	0	2	-2	-4	1	2	2	4
Maximize Retirement of Legacy Systems	0	2	2	4	-1	-2	-1	-2
				13		26		**28**

Table 1: Pugh matrix evaluating alternatives for information accuracy, usability, and cost

Section 7: Recommendations

Selecting the best course of action or suggesting what should be done is not something to be taken lightly. Determining the alternative that best meets the given objectives or one that provides the most benefit is only half of the battle. The other half is the analysis of risk associated with the decision to be made. Items to consider, such as past success and failures, opportunities of similar scale and customer criteria are all important. Past success and failure will talk to the risk associated with each decision. Opportunities of a similar scale will be used to

eliminate choices that do not have the potential to be a viable solution. Additionally, analysis of customer criteria or analysis of PPTI is highly recommended to ensure success of creating or even maturing an existing capability. In this analysis, recommendations in all four areas were made. However, the analysis focused on maturing the technology space. It is true; "adopting a broader, systems-oriented perspective should yield efficiencies by strengthening value-added intersections while eliminating efforts that are duplicative, ineffective, or irrelevant."[42] However, there are risks associated with every decision made.

To effectively design a future state, it is essential to evaluate past processes. Understanding what has not worked and under what circumstances is just as important as what has worked. For example, researching significant failures such as DIMHRS or the ECSS Project provides important guidance and a plausible avenue that should be avoided to be successful. DIMHRS and the ECSS Project are only but a few examples of failed attempts the government has made in commissioning data management systems reinforcing Santayana's famous dictum, "Those who cannot remember the past are condemned to repeat it."[43]

Risks such as utilizing a tool beyond its capability or designing a tool that does not address customers' objectives leads to a breakdown in capabilities that support PPTI. These problematic designs have cost the U.S. taxpayer billions of dollars, wasted a great deal of time and have negatively impacted those that try and use the systems provided. For example, DIMHRS was a system designed for use across all Services to improve human resource management for the military.

Unfortunately, after 12 years all it delivered was a non-viable product and a billion dollar debt. Sadly, the bigger problem was the conditions created that negatively affected "more than 90 percent of Army Reserve and Guard soldiers activated to serve in Afghanistan and Iraq

through 2003."[44] The use of an older programming language made it difficult to manage complex changes such as automating bonus programs. This difficulty led to the need of manually processing accounts, which increased the number of errors and delayed payments.

Another significant failure was the ECSS project. "The ECSS project began in 2004 as an ambitious and risky effort to replace 240 outdated Air Force computer systems with a single integrated system so that the Air Force could finally come up with an auditable set of financial records."[45] The ECSS project was terminated but still cost Taxpayers a billion dollars over seven years of development to produce a system which it admits as having no "significant military capability."[46]

According to the analysis completed above the alternative that provides the most value is a data hub system. Unlike ERP, scalability, which has been a significant problem with past designs, is not a problem. Through the use of code, it has the ability to communicate and potentially fully integrate with existing legacy systems leading to shorted delivery time. Similar to previously proposed systems, integrating various systems minimizes duplicative efforts and restores faith in data integrity due to "one version of the truth." On the other hand, the data hub system utilizes existing client apps also known as legacy systems to maintain information. This could be problematic due to the age and the operating cost of some the systems in place today.

Reviewing the risk from a potential problem perspective, several opportunities exist to address these issues. For example, in a civilian semiconductor manufacturing facility there was a need to develop a repository that enabled the use of relational type tables to improve automating of metrics, dashboards, and capacity analysis. Not wanting to give up the original mainframe due to the impressive reliability of the system, the company created a combination of two different technologies, data hubs and ERP, to actively synchronize data across different platforms. The

value gained was impressive. Not only was the company able to realize real-time updates in both systems but also little to no downtime was experienced on the manufacturing floor. Additionally, improved automation improved how the company was reacting to the volatile market leading to a substantial gain in market share.

In summary, history has proven traditional ERP systems will not work for the problem we are trying to address. Leaving existing systems in place is also not an option due to aging systems and cost of maintenance. While data hubs have proven to be the clear winner from a benefits perspective, the risk associated with leaving existing systems in place as a part of the architecture is far too high. For example, maintenance cost will continue to climb as the population with knowledge and experience of these legacy systems gets smaller. Manufactures are no longer making the same parts, making it difficult to replace parts that malfunction. However, there is an opportunity to leverage multiple technologies to accomplish what is needed for the Air Force and possibly all joint forces as a whole. Software is maturing and new code is being developed to accomplish more complex tasks much faster. Utilizing theories that have successfully been accomplished in the civilian sector, the Air Force can combine technologies in a way that would reduce the risk of implementing only one alternative, ultimately leading to a viable integrated solution that will be considered the real source of truth.

Endnotes

[1] Air Force Instruction (AFI) 20-210, *Operations: Prime Base Engineering Emergency Force (BEEF) Program*, 6 September 2012.

[2] Davidson, Lee, "Air Force Bases Are at Risk, Report Says," *Deseret News*, http://www.deseretnews.com/article/705295321/Air-Force-bases-are-at-risk-report-says.html?pg=all, 5 April 2009, accessed 12 Sept 2015.

[3] Kalil, Thomas, "Information Technology and Social Progress," http://www.huffingtonpost.com/thomas-kalil/information-technology-an_b_860582.html, 12 July 2011, Accessed 10 Jan 2015.

[4] Goldberg, Saveli, Niemierko, Andrzej, Turchin, Alexander, "Analysis in Data Errors in Clinical Research Databases," *US National Library of Medicine National Institutes of Health*, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2656002/, 2008.

[5] Ibid.

[6] Database Design Examples, Yeldiz Technical University, https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=0ahUKEwjh1M6n4tHKAhWMRiYKHWgJCAwQFggyMAY&url=http%3A%2F%2Fwww.yildiz.edu.tr%2F~kalipsiz%2FDatabase1.ppt&usg=AFQjCNFP6HDyj-UeX_lSOTb99LBhs4yvJQ, 3 March 2004.

[7] Weaver, Charles, "Does the Price of Technology Really Go Down Over Time?" http://mspalliance.com/does-the-price-of-technology-really-go-down-over-time/, 27 January 2012, accessed 20 January 2016.

[8] ACM Computing Surveys, "Report of the Working Group on Storage I/O Issues in Large-Scale Computing," http://www.cs.cmu.edu/~garth/SDCR/SDCR-final-report.pdf, December 1996, accessed 20 January 2016.

[9] Tilghman, Andrew, "Longer tours, fewer moves loom for troops: PCS cuts could shake up thousands of careers," *Military Times*, http://archive.militarytimes.com/article/20130708/NEWS/307080005/Longer-tours-fewer-moves-loom-troops, 8 July 2013, accessed 18 Jan 2015.

[10] Business Dictionary, "Problem," http://www.businessdictionary.com/definition/problem.html, accessed 19 January 2016.

[11] Redman, Thomas C., "Data's Credibility Problem," *Harvard Business Review*, December 2013, https://hbr.org/2013/12/datas-credibility-problem.

[12] Jagelski, Tina SMSgt, *149 FW Recall Performance Evaluation*, 12 March 2015.

[13] SolutionPipe, "Department of Defense-Military Air Force Directory Services," http://gov-spending.solutionpipe.com/investment/2152-2-air-force-directory-services, accessed 16 January 2016.

[14] Chen, Peggy, "Defense Manpower Data Center Wins Award for Excellence in the Workforce," Oracle WebCenter Blog, https://blogs.oracle.com/webcenter/entry/defense_manpower_data_center_wins, 06 April 2012, accessed 22 January 2016.

[15] DMDC, https://www.dmdc.osd.mil/appj/dwp/index.jsp, accessed 20 January 2016.

[16] Ibid.

[17] Ibid.

[18] Obamacare Facts, *ObamaCare Employer Mandate*, http://obamacarefacts.com/obamacare-employer-mandate/, 2016.

[19] Gildea, Debbie, "MilPDS Upgrade Begins," Air Force Personnel Center, http://www.afpc.af.mil/news/story.asp?id=123338605, 4 March 2013, accessed 18 January 2016.

[20] Air Force Instruction (AFI) 36-2134, *Personnel: Air Force Strength Accounting Duty Status Program*, 4 August 2014.

[21] Air Force Policy Document (AFPD) 38-2, "Manpower and Organization: Manpower," 5 February 2013.

[22] Defense Information System Agency (DISA), "Global Content Delivery Service (GCDS)," http://www.disa.mil/network-services/gcds, accessed 12 January 2016.

[23] *JOPES: Joint Planning and Execution System*, http://fas.org/nuke/guide/usa/c3i/jopes.htm, 2 June 1997.

[24] Power, Daniel J. "Ask Dan About DSS: What are some examples of database failures?" http://dssresources.com/faq/index.php?action=artikel&id=282, 18 August 2013, accessed 22 January 2016.

[25] Shaw, Katchina., "Lessons from a Billion-Dollar Project Failure," *ITBusinessedge*, http://www.itbusinessedge.com/blogs/governance-and-risk/lessons-from-a-billion-dollar-project-failure.html, 11 December 2012, accessed 22 January 2016.

[26] Business Process Definition, *Appian*, http://www.appian.com/about-bpm/definition-of-a-business-process/, 2016, accessed 1 February 2016.

[27] Change Management Process, Prosci.com, https://www.prosci.com/change-management/thought-leadership-library/change-management-process, 2016, accessed 2 February 2016.

[28] Leinwand Paul, "Only 8% of Leaders are Good at Both Strategy and Execution," *Harvard Business Review*, 30December 2015.

[29] Ibid.

[30] ERP Cloud: Streamline your enterprise business processes, Oracle Cloud, https://cloud.oracle.com/erp-cloud?tabname=CatCustomerInfo, accessed 3 February 2016.

[31] Manes, Anne Thomas, "Book Excert: When to Use Web Services," *Computer World*, http://www.computerworld.com/article/2566429/app-development/book-excerpt--when-to-use-web-services.html, 16 August 2004, accessed 22 February 2016.

[32] Ibid.

[33] J. Ronald, W. Christopher, C. Ralph, "An Interphase Transformation Strategy for AF-IPPS," *MITRE Corporation*, Bedford Massachusetts, 2012.

[34] "Top 5 Considerations When Evaluating NoSQL Databases," *MongoDB,* August 2015.

[35] Ibid.

[36] Press Release15-138, "Establishing a brain trust for data science," *National Science Foundation*, 2 November 2015.

[37] Bhardwaj, Anant, Bhattacherjee, Souvik, Chavan, Amit, "Data Hub: Collaborative Data Science & Dataset Version Management at Scale," MIT University, 2015.

[38] Ibid.

[39] E/M University, *Medical Decision Making*, http://emuniversity.com/MedicalDecision-Making.html, accessed 29 January 2016.

[40] Decision Making Confidence, "How to use the Pugh Matrix," http://www.decision-making-confidence.com/pugh-matrix.html, accessed 11 February 2016.

[41] Pratt, Mary, "CIOs make the tough call on legacy systems," CIO Innovators, http://searchcio.techtarget.com/feature/CIOs-make-the-tough-call-on-legacy-systems, accessed 19 February 2016.

[42] Conley, Raymond, "Maintaining the Balance Between Manpower, Skill Levels, and PERSTEMPO," *RAND Project Air Force*, http://www.rand.org/content/dam/rand/pubs/monographs/2006/RAND_MG492.pdf, 2006.

[43] Santayana, George, "Reason in Common Sense," *The Life of Reason,* p 284, 1905.

[44] Philpott, Tom, "Military Update: DIMHRS program dumped as 'a disaster'," *Stars and Stripes*, http://www.stripes.com/opinion/military-update-dimhrs-program-dumped-as-a-disaster-1.99558, 10 February 2010.

[45] Charette, Robert, "U.S. Air Force Blows $1 Billion on Failed ERP Project," IEEE Spectrum, http://spectrum.ieee.org/riskfactor/aerospace/military/us-air-force-blows-1-billion-on-failed-erp-project, 15 November 2012, accessed 12 February 2016.

[46] Ibid.

Bibliography

Air Force Instruction (AFI) 36-3803, *Personnel: Personnel Accountability in Conjunction with Natural Disasters or National Emergencies*, 30 June 2014.

Air Force Instruction (AFI) 33-332, *Communication and Information: Air Force Privacy and Civil Liberties Program*, 12 January 2015.

Air Force Manual (AFMAN) 33-363, *Management of Records: Communications and Information*, 9 April 2015.

Charette, Robert N., The U.S. Air Force Explains its $1 Billion ECSS Bonfire, http://spectrum.ieee.org/riskfactor/aerospace/military/the-us-air-force-explains-its-billion-ecss-bonfire, 6 Dec 2013, accessed 10 Jan 2016.

Davidson, Lee, "Air Force Bases Are at Risk, Report Says," *Deseret News*, http://www.deseretnews.com/article/705295321/Air-Force-bases-are-at-risk-report-says.html?pg=all, 5 April 2009, accessed 12 Sept 2015.

Inspector General United States Department of Defense, "DoD Generally Effective at Correcting Causes of Antideficiency Act Violations in Military Personnel Accounts, But Vulnerabilities Remain (Project No. D2012-D000FI-0118.000)," http://www.dodig.mil/pubs/documents/DODIG-2013-027.pdf, 12 November 2012.

Johnson, Nicole Blake, "DISA Estimates Millions in Data Center Savings," *FedTech*, http://www.fedtechmagazine.com/article/2014/07/disa-estimates-millions-data-center-savings, 18 July 2015, accessed 11 January 2016.

Kalil, Thomas, "Information Technology and Social Progress," http://www.huffingtonpost.com/thomas-kalil/information-technology-an_b_860582.html, 12 July 2011, accessed 10 January 2016.

Michael Gout (Lead Enterprise Information Technology Architect, USAA), interviewed by the author, 12 December 2015.

Redman, Thomas C., "Data's Credibility Problem," *Harvard Business Review*, December 2013, https://hbr.org/2013/12/datas-credibility-problem, accessed 19 August 2015.

SMSgt. Heather David-Damm (Force Support Superintendent, 149th FW, Texas Air National Guard), interviewed by author, 17 November 2015.

SMSgt. Tina Jagelski (Records Manager, 149th FW, Texas Air National Guard), interviewed by the author, 7 August 2015.